Data Processing and Interpretation

Thomas Ernst, Ph.D., and Linda Chang, M.D.

Department of Medicine, John A. Burns School of Medicine
University of Hawaii, Honolulu, HI

Introduction

The purpose of spectral processing in the clinical or research setting is to create clinically interpretable (qualitative) or quantitative data from raw magnetic resonance spectroscopy (MRS) data sets. The end result may be either in the form of a spectrum for visual interpretation, quantitative numbers, or a metabolite map in the case of spectroscopic imaging (SI).

Processing of MRS data generally differs from MRI data processing. The major reason for this difference is that MRS data contain signals from multiple metabolites, while typical MRI data sets involve only a single signal source of interest (water). Because the total signal strength of all metabolites in a given volume is of little use, MRS data sets have to be "deconvolved" to determine the signal strength of each individual metabolite of interest. With certain assumptions, the signal strengths can be converted into metabolite concentrations, which then provide the basis for spectral interpretation.

Data processing

From a technical point of view, the goal of spectral processing is to determine the signal strength of each metabolite in a given spectrum, which is proportional to the metabolite concentration. In the past, spectral analysis programs required extensive user interaction. However, some of the more recent automated programs allow fully automated analysis. While the quality of modern spectral analyses packages is generally high, it is advisable not to compare data that were processed with different packages, since the metabolite levels or ratios for a given data set will typically be inconsistent when processed with different packages.

Time-domain preprocessing

Spectral analysis can be performed in the "time domain", using raw data, or the "frequency domain", using "spectra" after Fourier transformation of the time domain data. Analysis in the frequency domain [1], i.e. the use of spectra, is more intuitive and will be discussed first. However, several preprocessing steps are typically performed in the time domain. The particular steps involved in frequency-domain spectral analysis are outlined below. These steps used to be performed sequentially and required manual guidance by a trained spectroscopist; however, more recent software packages are completely automatic and require little user interaction.

Time-domain preprocessing

Before raw data are transformed into the frequency domain (using the Fourier transformation, see below), spectral analysis typically involves several pre-processing steps that are performed in the time domain. These steps are designed to improve the appearance of the

processed spectrum and minimize the variability of the final metabolite levels. Typical preprocessing steps include:

- **Eddy-current correction:** The time-domain data (i.e. free induction decays or FIDs) are corrected to remove phase variations due to residual gradient-induced eddy currents [2, 3]. This correction utilizes eddy-current induced phase-variations in the water signal.
- **Removal of residual water signal:** The potentially large and interfering residual water signal is removed from the raw signal, using (digital) "notch-filters" that attenuate signals close to the water frequency. Of note, this step is commonly implemented after Fourier transformation, i.e. in frequency-domain [4].
- **"Apodization" or low-pass filtering:** Next, the raw data are subjected to a digital "low-pass filter". Low-pass filters involve multiplication of the data with a function that decays in time, such as a linear function or an exponential, to attenuate signals towards the end of the data acquisition window. This "apodization" step reduces the apparent noise in the spectrum, but at the cost of broader spectral lines.
- **Zero-filling:** Finally, the digital raw data are typically expanded in size, by adding zeroes at the end of the acquisition window (i.e. on the right side of the data). For example, the number of data points may be doubled from 1024 to 2048, etc. Zero-filling" increases the apparent digital resolution of spectra.

Fourier transformation

The preprocessed time domain data are then converted into a "spectrum", i.e. frequency domain data, by means of a Fourier transformation. The Fourier transformation determines for a measured raw data set the amount of signal at a given frequency. A spectrum represent a graph of the signal strength versus frequency after Fourier transformation.

Frequency domain preprocessing

Following Fourier transformation, several additional processing steps may be performed (at this point in the frequency domain).

- **Phase correction:** Metabolite peaks may have distorted line shapes after Fourier transformation. The reason is that the phase of the peaks may be incorrect. Manual or automatic programs make it possible to adjust the phase. Phase correction typically involves "zero-order" and "first-order" terms. For zero-order phase correction, the phase of all peaks is adjusted by the same amount. For first-order phase correction, the amount of phase correction applied is dependent (in a linear manner) on the frequency of each peak. First-order phase correction is required if the start of the data acquisition window is slightly shifted in time relatively to the optimum point. Note: Phase correction may be unnecessary if a time-domain eddy-current correction is used.
- **Baseline correction:** The "baseline" of phase-corrected spectra is commonly slanted or distorted, usually as a result of strong residual water or lipid signals. Manual baseline correction algorithms let the user define several spectral points as "baseline".

The computer fits a smooth curve through these “baseline” points and subtracts the fitted curve. Proper baseline correction may yield a spectrum with a much-improved baseline. Automated baseline correction algorithms may use a predefined set of “baseline” points instead of a manually-defined set. Other baseline-correction approaches include smoothly-varying functions during the fitting process (see next step). Note: Differences and inaccuracies in baseline correction algorithms probably explain a substantial amount of the variability in fitting results among spectral analysis programs.

Determination of metabolite peak areas

The preprocessing steps are designed to ensure optimal performance of the algorithms to determine metabolite peak areas, the final processing step. The peak area is proportional to the concentration of each metabolite, or the signal strength of the metabolite in the time domain. Of note, the peak area, but not peak height, represents metabolite concentrations.

Peak integration: One early approach for determining metabolite peak areas is by means of numerical integration, either manually or automatically. After selecting a frequency point to the left and one to the right of a peak of interest, the computer numerically integrates the area under the peak (with the assumption that the two frequency points reflect “zero” amplitude).

However, manual or automated peak integration has substantial limitations in the *in vivo* setting, for several reasons: 1) the exact boundary between peaks that resonate closely, such as that between total creatine and total choline, is often difficult to define. 2) Peak integration does not make it possible to separate signals from metabolites that co-resonate at the same frequency. For example, there is significant overlap between the peak for the N-Acetyl compounds (NA) at 2ppm, the multiplets from glutamate and glutamine between 2.0 and 2.4 ppm, and macromolecule resonances. As a result, peak integration will commonly result in areas that reflect an admixture of two or more peaks, and are very sensitive to changes in the spectral baseline.

Peak fitting: Many of the drawbacks of peak integration can be avoided by the use of iterative fitting algorithms. These algorithms model the experimental data as a superposition of ideal metabolite spectra. Fitting algorithms differ in the following features:

- *Fitting peaks independently:* Early algorithms fitted each metabolite peak separately. This can be sufficient for spectra with well-separated and well-defined resonances that show little overlap. For instance, long echo-time proton spectra of the brain can be approximated by just three individual, reasonably separated peaks (NA, total creatine, and total choline).
- *Incorporation of “prior knowledge”:* In the case of more complex spectra with multiple and overlapping peaks, fitting results are generally more robust and of higher quality when the fitting algorithm incorporates “prior knowledge”. Prior knowledge refers to spectral features that are known and do not vary from subject to subject, such as peak positions, or fixed intensity or phase relationships between peaks. For instance, it is advantageous to describe the lactate doublet as two related peaks that are separated by a certain chemical shift and show a fixed amplitude ratio. Similarly, the line widths of the different peaks in a given *in vivo* spectrum are not independent. Mathematically, prior knowledge reduces the number of parameters (or “degrees of

freedom") that need to be estimated by the fitting algorithm. The result is usually a more robust and more accurate fit.

- *Fitting model spectra:* Instead of describing *in vivo* spectra as a superposition of multiple single peaks of a given shape (possibly using prior knowledge), spectra can also be modeled as an admixture of ideal "basis" spectra that represent the major metabolites in the organ of interest. The basis spectra are acquired using model solutions of the metabolites of interest. For instance, ^{1}H MR spectra of the brain may be modeled by a "basis set" comprising spectra of N-acetyl-aspartate (NAA), creatine and phosphocreatine (CR), choline (CHO), myo-inositol (MI), glutamate, glutamine, lactate, gamma-amino-butyric acid (GABA), glucose, NAA-glutamate, etc. The contribution of each metabolite in the basis set is estimated by the fitting program, yielding the (relative) concentration of each metabolite in the measured spectrum. One example of this approach is the "LCModel" program [5].
- *Inclusion of lipid and macromolecule signals:* It has been demonstrated that short echo-time spectra of the brain (and other organs) contain contributions of macromolecules and lipids, especially in the 0.5 to 2.0 ppm region [6, 7]. These signals appear as relatively broad resonances, and may easily be confused with the "baseline". However, rather than modeling these signals with the "baseline", it is advantageous to include them as prior knowledge during the fitting process [8]. For instance, some of the LCModel basis data sets include typical macromolecule and lipid signals. Inclusion of these resonances generally improves the robustness and accuracy of the fitting algorithm, and results in a smoother baseline.
- *Time domain fitting:* This presentation has focused on fitting in the frequency domain, since it is relatively intuitive. However, frequency domain fitting has no intrinsic advantages over time domain fitting; in fact, the two are mathematically equivalent and both approaches are found in modern fitting programs.

Spectral interpretation

Interpretation of *in vivo* spectra in the clinical setting is different from interpretation of high-resolution structural images, in that spectral interpretation commonly relies on the use of quantitative information in the form of metabolite concentrations or ratios. Conversely, pure visual readings of spectral results are less common compared to MRI and should actually be avoided. The reason for this difference is that MRI relies on the detection of spatial abnormalities associated with disease conditions, whereas MRS interpretation commonly relies on the interpretation of differences in metabolite levels at a few locations. Furthermore, spectroscopic peaks reflect the concentrations of metabolites in the tissue, and it is impossible to determine these concentrations visually in a reliable and accurate manner. Nevertheless, it is important to examine each spectrum visually, in order to ensure that no artifacts are present and that the result of the fit is accurate.

The following steps have proven useful for the interpretation of single-voxel data. MRS analysis as described above will usually yield a spectrum as well as quantitative peak information in the form of metabolite concentrations or ratios.

- An important first step is to visually inspect each spectrum. Visual inspection will allow an experienced reader to evaluate the quality of the spectrum, as well as

identify potential artifacts. Spectra with major artifacts need to be interpreted with caution.

- Ideally, the spectral analysis package will display the raw spectrum after processing, as well as the fitted result as an overlay. Comparison of the fitted versus the measured spectrum makes it possible to assess the quality of the fit, and consequently the quality of the estimated metabolite levels.
- An experienced reader might also be able to draw some conclusions with regards to possible spectral abnormalities, especially if the abnormalities are substantial. For instance, a dramatically reduced NA to creatine ratio may be indicative of neuronal loss. However, it is virtually impossible to evaluate metabolite concentrations by means of visual interpretation; only metabolite ratios can be assessed visually.
- It is advisable to evaluate the quantitative results of the fitting algorithm, and compare them with the visual interpretation. For instance, an apparently low NA to creatine ratio during visual inspection should be reflected in a low NA to creatine ratio from the fitting algorithm.
- Ideally, fitting algorithms yield metabolite "ratios", for instance the ratio of NA / creatine, as well as metabolite concentrations. Importantly, interpretation of metabolite ratios is inherently ambiguous, since it is impossible to determine from metabolite ratios whether the nominator or denominator metabolite (or both) are abnormal. Therefore, it is preferable to interpret metabolite concentrations, and not metabolite ratios.
- For the interpretation of more subtle spectral abnormalities, it is important to know the normative values of each metabolite, for instance, in the form of a normal range or the mean value with its standard deviation. Individual metabolite values (concentrations or ratios) should be considered abnormal only if they are 2 standard deviations above or below the normal value. Thus, metabolite levels should be interpreted similar to routine clinical blood chemistry results.
- It is important to emphasize that normative values for metabolites generally vary by region (e.g. gray matter vs. white matter) as well as by age and sex. For instance, the concentration of NAA is higher in gray matter than in white matter, while choline compounds are higher in male than female brains. The concentrations of creatine and myo-inositol generally increase during normal aging. Therefore, regional, age and sex-appropriate normative data are needed for comparison and interpretation of the spectral data.
- Interpretation of metabolite concentrations also may need consideration of partial volume effects. For instance, cerebrospinal fluid (CSF) contains very low concentrations of the major brain metabolites; therefore, metabolite concentrations are ideally corrected for the presence of CSF during spectral analysis.

Interpretation of spectroscopic imaging (SI) data sets is somewhat more reliant on the evaluation of the spatial distribution of metabolites. While the interpretation of each metabolite map depends on the specific clinical question, the following hints may be helpful.

- If a metabolite map reveals potential abnormalities, it is advisable to confirm these abnormalities on the "raw" spectrum. This is important since 1) there are many potential sources of artifacts, including poor water suppression, distorted baseline, excessive line-width, residual lipid signals, and motion artifacts, and 2) artifacts in low-resolution SI scans/images are generally much less obvious compared to artifacts in high-resolution structural MRI scans.
- If the SI processing package provides maps of metabolite concentrations as well as ratios, it is generally better to evaluate metabolite concentrations instead of ratios (see above).

References

1. Mierisova, S. and M. Ala-Korpela, *MR spectroscopy quantitation: a review of frequency domain methods.* NMR in Biomedicine, 2001. **14**(4): p. 247-59.
2. Klose, U., *In vivo proton spectroscopy in presence of eddy currents.* Magnetic Resonance in Medicine, 1990. **14**(1): p. 26-30.
3. Lin, C., et al., *Eddy current correction in volume-localized MR spectroscopy.* Journal of Magnetic Resonance Imaging, 1994. **4**(6): p. 823-827.
4. Coron, A., et al., *The filtering approach to solvent peak suppression in MRS: a critical review.* Journal of Magnetic Resonance, 2001. **152**(1): p. 26-40.
5. Provencher, S., *Estimation of metabolite concentrations from localized in vivo proton NMR spectra.* Magnetic Resonance in Medicine, 1993. **30**: p. 672.
6. Behar, K., et al., *Analysis of macromolecule resonances in ^{1}H NMR spectra of human brain.* Magnetic Resonance in Medicine, 1994. **32**: p. 294-302.
7. Seeger, U., et al., *Elimination of residual lipid contamination in single volume proton MR spectra of human brain.* Magn Reson Imaging, 1999. **17**(8): p. 1219-26.
8. Seeger, U., et al., *Parameterized evaluation of macromolecules and lipids in proton MR spectroscopy of brain diseases.* Magn Reson Med, 2003. **49**(1): p. 19-28.